

The ESA Data Warehouse: A Tool for Effective Management Reporting

Y-M. Bourguignon, M. Chevrier, M. Douzal & G. Gendreau

Information Systems Department, ESA Directorate of Administration,
ESRIN, Frascati, Italy

Introduction

Management information relies on well-organised, easily exploitable data. As managers need more information, their data reporting and analysis requirements become more stringent. At the same time, traditional information systems are often not in a position to deliver the information required, chiefly because they specialise in one functional area – and have as a result a limited data scope – and because their functional focus is on data entry and processing, not reporting. As their reporting and analysis needs outgrow the capabilities of traditional information systems, therefore, organisations like ESA are turning to Data Warehouses to fulfil their needs.

A data warehouse is a data reporting and analysis system that, thanks to its specific functional and architectural features, provides uniform and flexible access to a wide range of data. The ESA Data Warehouse project was brought about by the combination of growing requirements from administrators and managers, and the limitations of the data reporting and analysis capabilities of traditional information systems. The system has been made available incrementally and is now being used extensively by all Directorates. The project as originally defined is now in the final stages of implementation.

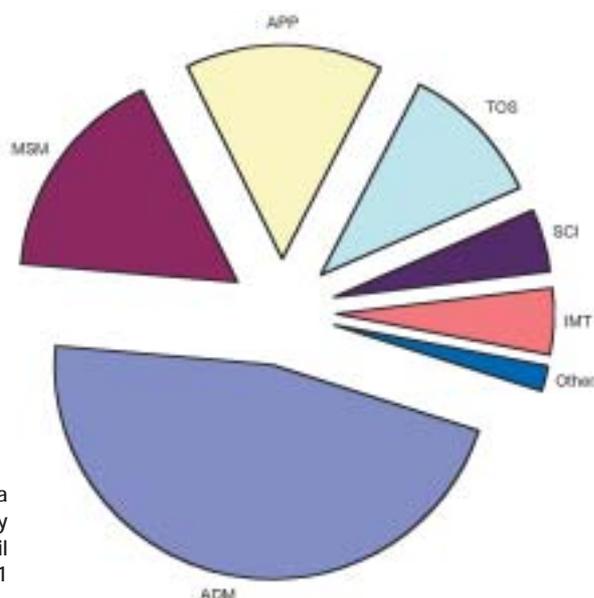


Figure 1. ESA Data Warehouse usage by Directorate, status April 2001

The major features and benefits of the ESA Data Warehouse fall into three broad categories:

- cost and operational aspects
- data integration aspects, and
- the time dimension.

The Warehouse's functionality has been made available incrementally so that value would be delivered to users as it became available. From the precursor systems to the full deployment by the end of 2001, the system has been increasingly popular for both administrative and management tasks (Fig. 1). Now that the initial project is nearing completion, the Warehouse is offering further opportunities, taking advantage of the structure, data and tools readily available.

The problem with reporting

Despite their name, Management Information Systems (MIS) often fall short of providing managers with the full supply of information they require. The reasons for this are linked to the very purpose of traditional information systems and fall into three categories:

- A MIS is created to fill a specific, operational data-processing need in some organisational process. It must focus on the process, the corresponding data and related calculations. Data outside the specific functional scope of the system is handled only to the limited extent of the process interfaces. A MIS is therefore a specialised system that handles a limited scope of data.
- The functional focus of a MIS is on the processing to be performed, which often includes data entry by the end-users. This leads to technical choices in which reporting needs cannot be the main driver. For example, the data structure will normally be optimised for responsiveness in data entry at the expense of reporting. When necessary, older data no longer needed for day-to-day transactions, but still useful for reporting, will be purged to make room for new data.

- Even if the combined MIS available could cover the data span required for management, it would be a frightening prospect to have to master all of the different systems, with their idiosyncrasies, access limitations, licence costs, etc., when all you need is to extract information.

On the other hand, management requires an integrated view of management data, with a single reporting environment and powerful analysis tools spanning the whole spectrum of data, from finance to contracts to programme management, and from past to current to planned. A data-warehouse system has specific features that address the reporting problem.

Features and benefits

A data-warehouse system specialises in collecting, storing and rendering administrative and management data. Manual data entry or transaction processing are deliberately beyond the scope of a data warehouse: the data is collected from existing sources like transactional systems or archives – obviously after formal agreement between the data owners and the users of the data warehouse. The data is refreshed in principle once a day, at the end of the working day to avoid any impact of the refresh process on the response time of the data sources.

The Conceptual Data Model (CDM) of the data warehouse represents the data entities handled in the warehouse and their relationships. The representation of each data entity is unique, follows a single set of standards and conventions, and brings together all the data attributes and relationships regardless of their source system.

The physical database of the data warehouse reflects CDM closely, with some redundancies added to optimise rendering, typically in the form of pre-calculated aggregations. This is a major difference between a data warehouse and a transactional, data-entry-oriented system: the former is optimised for providing information *to* the end user, the latter for obtaining and processing information *from* the end user. The result is a single database, structured to reflect functional constructs, and seamlessly merging data from a variety of source systems.

At the user end, the data is made available to end users and client applications – within the limits prescribed by the original owners of the data. The client applications fetch data as input to their own operations. The end users exploit the data by means of the various query,

reporting and analysis tools, collectively known as 'IRMA'.

The most notable of those tools is Business Objects (BO), a market leader among so-called 'business intelligence tools', which has been introduced to allow users without knowledge of the technical data structure to perform ad-hoc queries on the data, and such analyses as 'drill-downs' from aggregate into detailed information.

This architecture leads to three families of features and benefits, related to:

- data integration
- time dimension
- operations and costs.

Data integration

Classical information systems automate the processes of one business unit: within ESA, these are AWARDS for Finance, COSY for Contracts, IPMS for Personnel, etc. To achieve this, they store and manage data of specific relevance to the business process they support: contracts, budgets, and staff information. They offer reporting functions centred on this business process data. This creates a double driver for data integration.

The first driver comes from the need, in every system, to handle data of ESA-wide relevance, like lists of Outputs, Directorates and General Headings – and data produced by other systems: industrial-return calculations in SYSTRI are based on commitments created in AWARDS.

The natural solution – building peer-to-peer interfaces – is costly to put in place and to maintain because of the number of interfaces required. If a system receives data from p providers and supplies data to r recipients, the interface cost for this particular system in a peer-to-peer approach is $p+r$.

The Operational Data Store (ODS) is the part of the DW that acts as a broker between the partner systems. On the basis of Data Exchange Agreements established between provider and recipient, ODS collects, integrates and stores the data from one and makes it available to the other. No matter how many providers and recipients there are, the cost to each individual system is reduced to one or at most two interfaces.

The second driver for data integration comes from those users who need broader data visibility than is provided by a single system. The data warehouse caters for these users, allowing, for example, seamless navigation between financial data and related contractual

data: the user is not aware that the data originates in different systems.

Time dimension

Retaining a memory of the past is necessary for various reasons. Not only does the legal status of ESA require that at least the last 8 years' worth of financial data be available for on-line consultation, but there are also many other functional uses for historical data.

The term historical – as opposed to operational – refers to data that is no longer kept in its original transactional system, either because it was purged to make room for fresh data (as happens in AWARDS), or because the original system is no longer in operation. The latter is notably the case of EFSY, the predecessor of AWARDS as ESA's financial management system.

The data model of the data warehouse reflects the functional concepts of the transactional systems, but not the technical idiosyncrasies of these systems. This has made it possible (if not easy) to collect, structure and store together data coming from EFSY and AWARDS, which are functionally similar but technically totally different systems. Furthermore, the data warehouse is designed in a way that allows many years' worth of data to be stored and retrieved without penalising performance.

There is more to handling historical data than just accumulating records. It is one thing to accumulate many individual records, created over many years and never changing, like transaction records. It is another to handle information which evolves with time, like the status of commitments. For this, it was necessary to equip the system with a 'snapshot' function able to freeze the status at any given time. Depending upon the type of information handled, the granularity retrievable from the data warehouse ranges from a month to a year. The time interval between snapshots is an adjustable system parameter.

A further dimension to the problem of handling historical data is that the referential data changes over time. These are the pointers used to classify and retrieve the data: lists of Directorates, Establishments, Outputs, etc. After a reorganisation, you need to be able to retrieve the data both through its original reference (former Directorate) and through the new one.

By combining these capabilities, the data-warehouse user can run the exact same queries and reports that are available on current data on data dating back to 1988.

These standard queries and reports bear upon the data of one particular year, but if requested an analysis of trends spanning many years could be implemented on the same basis.

Operational and cost-related benefits

Having no data entry makes it possible to optimise the data structures of the data warehouse for efficiency in querying and reporting rather than data entry and transaction processing, which are the main drivers in designing transactional information systems. This is how complex reports can be obtained in the data warehouse with a response time that could not be matched by a transactional system.

Distributing the data entry and the reporting functions over different systems offers the additional benefit that the load can be balanced and the respective systems sized much more accurately. The population of users involved in data entry is a fraction of the population of report users: why bear the costs of a transactional system sized to support both?

Moreover, the separation of the reporting and data-entry functions provides many read-only users with access to reporting functions that they would otherwise not exploit, particularly when the transactional system is a commercial product. Quite a proportion of report-only users would be deterred by the cost of the necessary user's licence, the time needed to learn the idiosyncrasies of the system, or lack of the necessary privilege to use the transactional system.

Finally, the data structure and technical environment of transactional systems can be very complex, which requires development and maintenance personnel with specialist technical knowledge of the transactional system in question and makes it very costly to develop and maintain pre-defined queries and reports under diverse and complex environments. The data warehouse smoothes out the development and maintenance costs by operating in a standard technical environment, and only with the subset of the data structure that has functional relevance.

Value to ESA

The value of a data warehouse to an organisation is measured by the data available from it and the usage that is made of its services within the organisation. The ESA Data Warehouse handles about 150 different data entities, not counting ancillary or technical data. The domains covered (and corresponding partner systems supplying or reading data) currently include:

- finance (AWARDS)
- contracts (COSY)
- missions (MOS)
- personnel, excluding confidential data (IPMS)
- geographical return (SYSTRI)
- invoicing (EFIS).

Table 1. ESA Data Warehouse usage statistics

Number of IRMA active users: 257

Number of Business Objects/WebIntelligence IRMA universe users: 73

Number of FRS Financial reports produced

	May 2001	June 2001	July 2001
Total number	3982	4090	4045
Average per working day	181	195	184

The 5 most used FRS financial reports

Total number of instances produced from April 1999 to August 2001:

Accounting Situation of Expense Lines (582)	26 411
Current Status of Obligations (567)	26 374
Budget Summary (566)	25 546
List of Invoice Lines (574)	18 201
Expenditure by Directorate, Establ. and EPA (564)	2 045

The 5 most used tables in the Corporate Data Browser (BAR)

Total number of calls from December 2000 to date:

Commitments	7905
ESA Staff Missions	6564
Invoices	4958
Budget by Item	4565
Allocations	1698

Programme management systems are gradually appearing. AMS (TOS), AIMS (APP) and IPCDB (MSM) are already retrieving data from the Data Warehouse and have taken steps towards supplying data as well.

Figure 2 shows the data flows currently supported by the ESA Data Warehouse. The user community is comprised of end-users and application owners. The blend of pre-defined reports and guided queries and free-hand analysis tools helps hundreds of users with their administrative and managerial information needs. Table 1 presents some usage statistics. Figures 3a,b show examples of pre-defined report and user-driven data analysis that can be obtained from the same environment.

The two salient examples of AWARDS and SYSTRI show how applications can derive value from the Data Warehouse. AWARDS has delegated most of its reporting and interfaces to it. As a result, it was able to concentrate more on its core function of managing ESA's financial processes. And as its data base was threatening to become clogged up with data still required for reporting but not for day-to-day operations, it also archived 2 years' worth of data into the Data Warehouse and thereby avoided performance degradation. The Warehouse seamlessly absorbed this data and no difference was visible to the end-users.

SYSTRI, the system used by ESA's Industrial Policy Office (IPO) to calculate the industrial

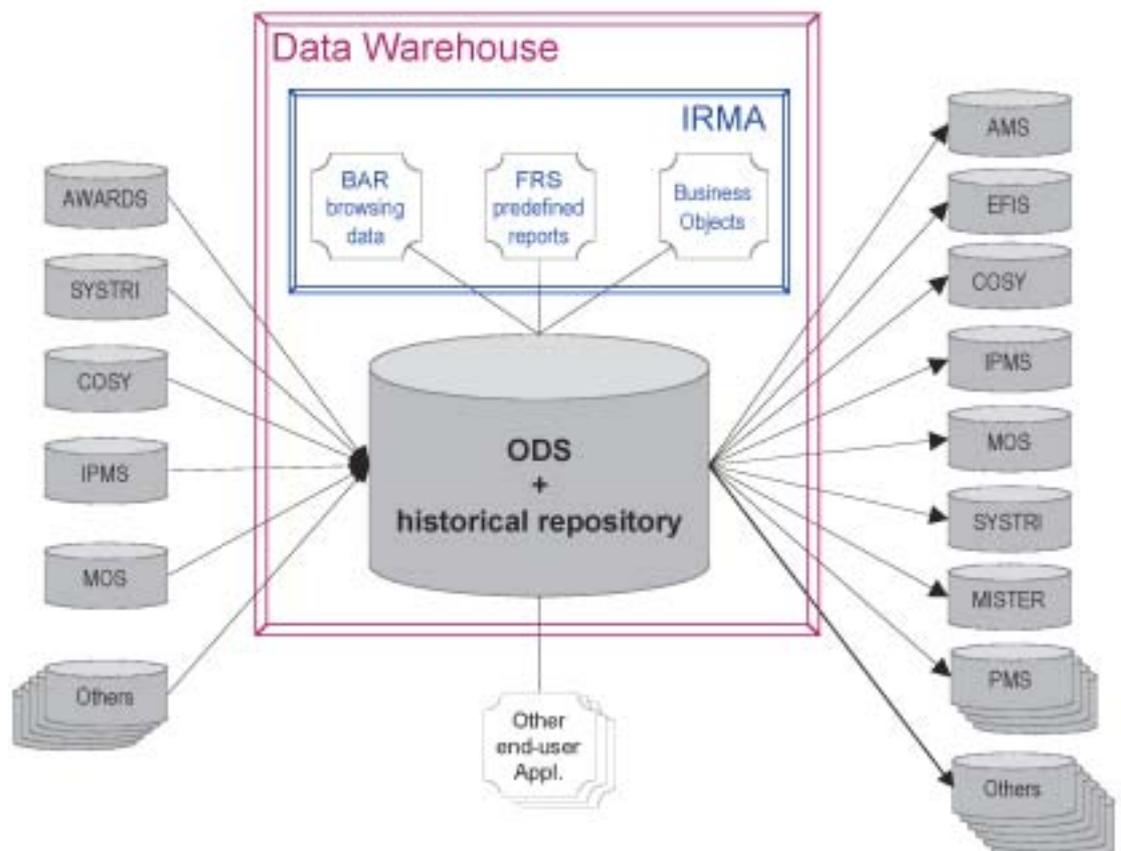


Figure 2. Data flows currently supported by the ESA Data Warehouse

return to Member States from the Agency, relies on data entered by end-users in IPO and in the Programmes, and on a large amount of data centred on the financial commitments. A lot of the basic data handled in SYSTR1 are also in the Data Warehouse, and require a similar structuring and archiving approach. This has allowed a high degree of synergy and integration between the two systems, reflected by the commonality of large parts of the conceptual data models as well as the physical data and the use of the same user environment and user tools, in particular Business Objects.

The next wave of project-controller-oriented functionality in SYSTR1, including the management of Work Breakdown Structure data below the financial output level, will be a new opportunity to share data, techniques and tools between the two systems.

Intangible benefits

In addition to these tangible benefits, the introduction of a data warehouse, through the use of a corporate conceptual data model as its foundation:

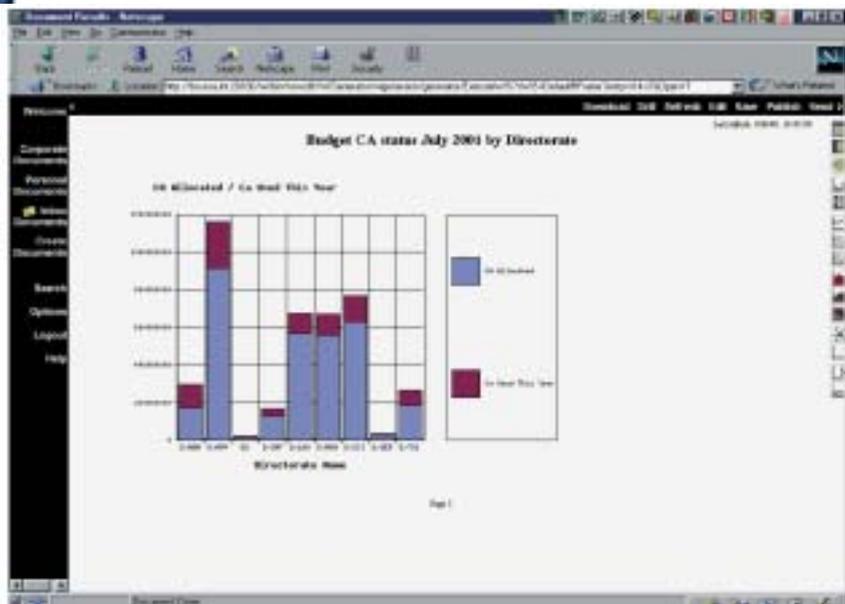
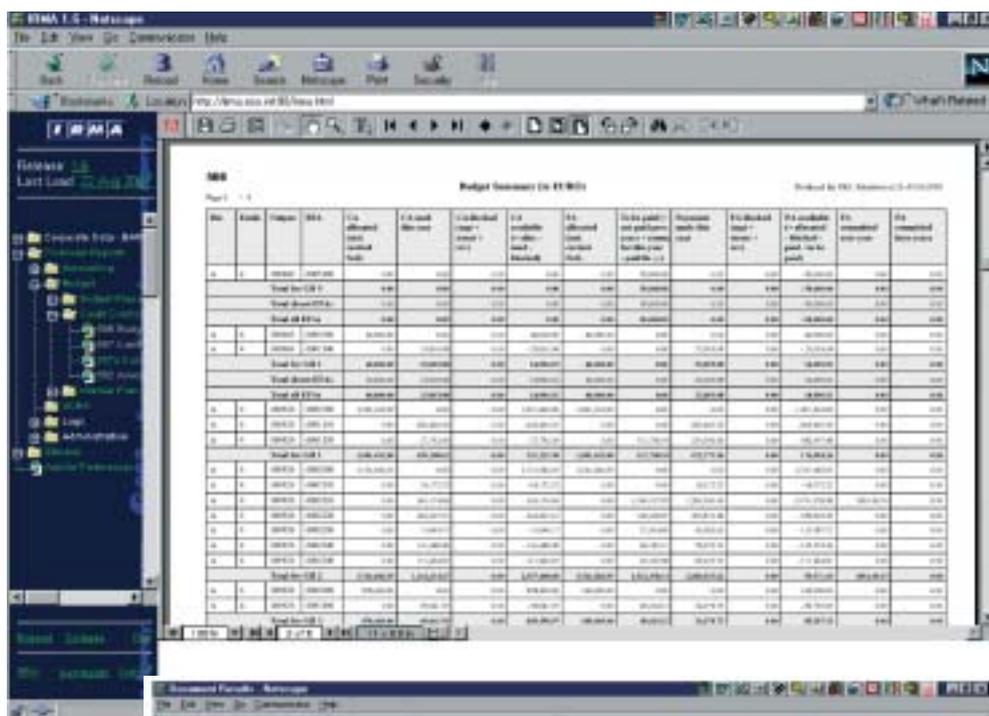
- fosters the use of a common language throughout the organisation
- facilitates the circulation and common understanding of information, and
- minimises the duplication of inconsistent data repositories.

Having hundreds of users from all Directorates use information based on one logical representation of ESA's data structure promotes a common language and effective communication. It also reduces ambiguity in information exchange and learning curves in cases of staff mobility.

There is still a lot to do in this domain, as can be seen from the multiplicity of interpretations of even such basic and ESA-wide concepts as the financial 'output'. This can be achieved by complementing the corporate data model with a user-readable data dictionary, which is part of the more general 'meta-data management'. The corporate data model, which is the foundation of the Data Warehouse, is also a major enabler of further evolution.

Evolution

The ESA Data Warehouse as it was initially defined will be complete by the end of 2001, with the implementation of outgoing



incremental interfaces for the partner systems reading data out of ODS, and the addition of features to enhance availability. Because of the Warehouse's careful design, enlarging its coverage to include additional data types and providing the corresponding capabilities for reporting and analysis comes at an ever-decreasing marginal cost. Now that the major referential entities (financial outputs, Directorates, OBS, EPA, etc.) are in place, adding new data is like adding a piece to an existing jigsaw puzzle. Business Objects, originally a powerful but novel tool, is now past its 'running in' stage, and the roles and procedures around the exchange of data through ODS are starting to be well understood.

Figures 3a,b. Examples of a pre-defined report and a user-driven data analysis obtained from the same environment

Of course, further evolution will be dictated by demand. Two major fields of application are provided by high-level reporting in Executive

Information Systems (EIS) and by the management of the full life cycle of obligations.

Executive Information Systems

Executive Information Systems (EIS) is a generic name for reporting systems that rely on large amounts of data and provide an aggregated and summarised view of it for use by top management. Since ESA's core business is in its programmes, an EIS that would:

- bring together planned data from Programmes
- integrate it with actual and past data from administrative systems, and
- provide reporting and analysis capabilities

could provide a consolidated view of ESA's situation and outlook with respect to payments, commitments, or other parameters.

The challenge is in reconciling the various data structures used in the various Directorates. The concepts used in the central system must be 'compatible' with those used in the individual Directorates and their local management systems, and with the Data Warehouse itself: there must be an unambiguous way of relating the data together. For example, while every programme should remain free to the concepts of project or activity at its convenience, a standard level must be defined in relation to the hierarchy of financial outputs for reporting purposes.

Management of obligation data throughout its life cycle

Before an activity formally becomes a row in the AWARDS commitment database, it follows a period of gestation as a planned activity. It then takes the shape of a contract, translating into commitments, invoices and payments according to a milestone payment plan. The stages of this life cycle are supported by different specialised systems in ESA: various local management tools (LMT) are used in the planning phase (and beyond), COSY handles contracts, AWARDS financial commitments, and EFIS receives milestone payment plans, invoices and payments.

Two major functions would help the end-to-end management of obligations in ESA. The first is the tracking of the obligation through its various stages of planning, contracting and executing. This would allow following the various stages of the 'activity', with the whole history available from one user environment. That environment could act as a portal, opening an access to specialised systems like COSY or EFIS to see more detail or take action on an obligation. A corollary function would aggregate the amounts 'accumulated' at the various stages of the life cycle of the activities.

Another major function will come out of SYSTRIS users' initiative to link the obligations to a breakdown of work packages below the level of the outputs. The user group is in the process of defining the required functionality. Depending on the requirements, other structures can be envisaged in addition to the WBS as axes of analysis: Product Tree, Organisation Breakdown Structure, Cost Breakdown Structure, etc. Any structure with at least its top level defined at corporate level is a candidate axis for aggregation and analysis.

As for EIS, further endeavours are contingent on a degree of coordination, so that:

- regardless of local differences in management processes, a number of well-defined aggregating concepts should be used identically throughout ESA; e.g. each programme should be free to define its WBS as it sees fit, but the concept of a work package should have one corporate meaning
- objects that users need to track throughout their life cycle should be assigned unique identifiers at the start of the life cycle and these should never be reused.

Conclusion

Like many organisations, ESA has equipped itself with a Data Warehouse to address its growing needs in the reporting and analysis of administrative and management data. The development originally planned has reached its final stages. The user functions, reporting, analysis and interfacing have been made operational gradually throughout the development and are already very popular. Additional value lies in the common language and improved communication brought about when data structures of corporate interest are standardised ESA-wide.

The data-warehouse concept provides a guarantee of maintaining the historical memory of an essential ESA management data set in periods of change in management information systems, ensuring proper access to the data in a coherent and understandable way. In addition to its current value, the data warehouse offers a wealth of well-structured data that serves as a platform for evolutions whenever required. Of particular interest is the possibility of aggregating programme-management data from the Directorates. This function can yield powerful results if a simple approach to standardisation is followed.

Further information on the features of and access to the ESA Data Warehouse can be obtained from the ESRIN Help Desk. 